deepest = "most protected",
deep = central, median = point with maximal depth (Tukey median)

1. Half-space / Tukey depth → computationally expensive (∞ planes for each point)
2. Simplicial / Liu depth → soldiers & queen, still computationally heavy
(3. Oja depth) ← <u>not</u> invariant w.r.t. translations
4. Mahalanobis depth → M. distance of one point w.r.t. the others
   → mean & variance <u>not</u> robust ⇒ method <u>not</u> robust
5. Convex-Hull peeling depth

$\underline{x}_1, \dots, \underline{x}_n$ points ⟶ $\underline{x}_{[1]}, \dots, \underline{x}_{[n]}$ points : $\underline{x}_{[1]}$ = deepest point
$\underline{x}_{[n]}$ = most outlying point
larger rank is associated with a more outlying position

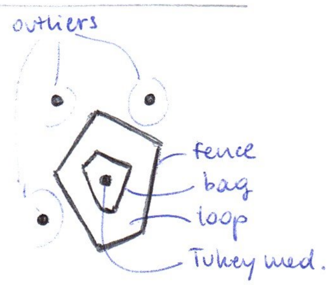
---

p-th central region = region containing at least p% of the density of the distribution

---

Graphical representations (generalizations of the box-plot)

1. Convex-hull → convex layers
2. Bagplot → bag ⇒ contains 50% of data (the deepest)
   Tukey median ⇒ obs. with max depth ($\in$ bag)
   Fence ⇒ separates the outliers from the other obs.
   loop → data which are not outlier but $\notin$ bag
   → effective up to 3 dimensions



3. Sunburst → bagplot-simile (central bag: 50% data (deepest))
4. DDplot → useful for comparison of different samples

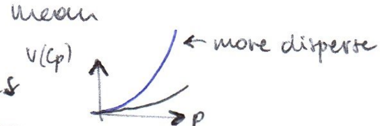
Measure of robustness:

1. breakdown point → percentage of points needed to make the estimator very different
2. Empirical influence function → measures how much an estimator depends on the sample

LOCATION:
Tukey median, simplicial median, & Co. are unbiased estimators for the <u>mean</u>.

DISPERSION:

- variance-covariance mtx with the median instead of the mean (the determinant is the "generalized sample scale")
- (p% contained density) vs. (volume of C_p region) as p varies



IC PARAMETRICI – One population

$X_1,\dots,X_n \sim N(\mu,\sigma^2)$, σ^2 known	$IC(\mu) = \left[\bar{x}_n \pm \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right]$, $IC(\sigma^2) = \left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right]$
$X_1,\dots,X_n \sim N(\mu,\sigma^2)$, σ^2 unknown $\frac{\bar{X}_n-\mu}{S/\sqrt{n}} \sim t(n-1)$, $S^2 := \frac{\Sigma(x_i-\bar{x}_n)^2}{n-1}$	$IC(\mu) = \left[\bar{x}_n \pm \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1)\right]$
$X_1,\dots,X_n \sim Be(p)$	• Conservativo ($\sqrt{p(1-p)} < 1/2$): ($p \sim \frac{1}{2}$) $IC(p) = \left[\bar{x}_n \pm \frac{1}{2\sqrt{n}} z_{1-\frac{\alpha}{2}}\right]$ • Semplice: $IC(p) = \left[\bar{x}_n \pm \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}\, z_{1-\frac{\alpha}{2}}\right]$ • Aggiustato: ($\bar{x}_n^* = (\Sigma x_i + 2)/(n+4)$) $IC(p) = \left[\bar{x}_n^* \pm \sqrt{\frac{\bar{x}_n^*(1-\bar{x}_n^*)}{n}}\, z_{1-\frac{\alpha}{2}}\right]$

IC parametrici – Two populations

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

$X_1,\dots,X_n \sim N(\mu_1,\sigma_1^2)$, σ_1^2 known $Y_1,\dots,Y_m \sim N(\mu_2,\sigma_2^2)$, σ_2^2 known	$IC(\mu_1-\mu_2) = \left[\bar{X}_n - \bar{Y}_m \pm \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\, z_{1-\frac{\alpha}{2}}\right]$
$X_1,\dots,X_n \sim N(\mu_1,\sigma^2)$, σ^2 unknown (but same) $Y_1,\dots,Y_m \sim N(\mu_2,\sigma^2)$	$IC(\mu_1-\mu_2) = \left[\bar{X}_n - \bar{Y}_m \pm t_{1-\frac{\alpha}{2}}(n+m-2)\, S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\right]$
$X_1,\dots,X_n \sim N(\mu_1,\sigma_1^2)$ σ_i^2 unknown $Y_1,\dots,Y_m \sim N(\mu_2,\sigma_2^2)$	$IC(\mu_1-\mu_2) = \left[\bar{X}_n - \bar{Y}_m \pm t_{1-\frac{\alpha}{2}}(\gamma)\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}\right]$, $\gamma = \dfrac{(S_1^2/n + S_2^2/m)^2}{\frac{(S_1^2/n)^2}{n-1} + \frac{(S_2^2/m)^2}{m-1}}$
$X_1,\dots,X_n \sim N(\mu_1,\sigma_1^2)$ $Y_1,\dots,Y_m \sim N(\mu_2,\sigma_2^2)$	$IC\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = \left[\frac{S_1^2}{S_2^2} f_{\frac{\alpha}{2}}(m-1,n-1) \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2} f_{1-\frac{\alpha}{2}}(m-1,n-1)\right]$
$X_1,\dots,X_n \sim Be(p_1)$ $Y_1,\dots,Y_m \sim Be(p_2)$	$IC(p_1-p_2) = \left[\bar{X}_n - \bar{Y}_m \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n} + \frac{\bar{Y}_m(1-\bar{Y}_m)}{m}}\right]$
Paired data: $D_i = X_i - Y_i$ $D_1,\dots,D_n \sim N(\mu_D,\sigma_D^2)$	$IC(\mu_D) = \left[\bar{D}_n \pm t_{1-\frac{\alpha}{2}}(n-1)\,\frac{S}{\sqrt{n}}\right]$

Exactness of permutations

Permutations are more powerfull than rank but less robust

What is the power, the control of errors (type I, II) etc

DEPTH MEASURES

Goal: rank the data (from the more centered to the more outlying):

The green points are deeper than the red ones.

We need a depth measure to rank them (= find their deep-ness)